83

AF/2654

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

NCR Docket No. 8999

Application of:

ARROWOOD, J. A. et al.

Group Art Unit: 2654

Serial No. 09/731,084

Examiner: Kinari M. Patel

Filed: December 6, 2000

For: NOISE SUPPRESSION IN BEAM-STEERED MICROPHONE ARRAY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF TRANSMITTAL LETTER

Sir:

Transmitted herewith for filing is an Appeal Brief and two copies thereof to the Final Rejection dated February 11, 2004.

☒ Please charge Deposit Account No. 50-1673 for the Appeal Brief fee or any other fees associated with the filing of said Appeal Brief.

☒ Please charge any additional fees to the account of NCR Corporation, Deposit Account No. 50-1673.

Respectfully submitted,

James M. Stover
Reg. No. 32,759

NCR Corporation
Dayton, Ohio
Tel. No. (937) 445-7663
Fax No. (937) 445-4792

CERTIFICATION OF MAILING UNDER 37 CFR 1.8

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on July 12, 2004

By
Name: JAMES M. STOVER

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Dayton, Ohio

Docket No. 8999

Application of:

Jon A. Arrowood et al.

Group Art Unit: 2654

Serial No. 09/731,084

Examiner: Kinari M. Patel

Filed: December 06, 2000

For: NOISE SUPPRESSION IN BEAM-STEERED MICROPHONE ARRAY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

BRIEF ON APPEAL

Sir:

This is an appeal under 37 CFR 1.191 to the Board of Patent Appeals and Interferences of the United States Patent and Trademark Office from the final rejection of claims 1 through 6 of the above-identified patent application. The claims were finally rejected in an Office Action dated February 11, 2004. Three copies of the brief are filed herewith, together with the requisite fee under 37 CFR 1.17(f).

07/15/2004 MBLANCO 00000075 501673 09731084
01 FC:1402 330.00 DA

CERTIFICATION OF MAILING UNDER 37 CFR 1.8

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on July 12, 2004.

By: [signature]
Name: James M. Stover

(1) REAL PARTY IN INTEREST

The present application is assigned to NCR Corporation.

(2) RELATED APPEALS AND INTERFERENCES

There are currently no known active appeals or interferences related to the present application.

(3) STATUS OF CLAIMS

Claims 1 through 6 are pending in the application.

Claims 1 and 2 are rejected as being unpatentable over Martino et al (US Patent No. 6,061,646) in view of Slyh et al. (US Patent No. 5,574,824). Claims 3 and 4 are rejected as being unpatentable over Martino et al. in view of Slyh et al. and Nagata (US Patent No. 6,009,396). Claims 5 and 6 are rejected as being unpatentable over Martino et al. in view of Nagata The rejections of claims 3 through 6 are being appealed.

The claims are shown in the Appendix attached to this Appeal Brief

(4) STATUS OF AMENDMENTS

A response to the Final Rejection dated February 11, 2004 and canceling claims 1 and 2 was filed on May 10, 2004. The response has not been entered.

(5) SUMMARY OF INVENTION

The present application describes and claims an improved beam-steered microphone array which can be utilized within telephone conferencing systems, automated teller machines (ATMs) and self service kiosks, bank and restaurant drive-up windows. The beam-steered microphone array includes electronic circuitry that controls the selection or orientation of one or more microphones, or

lobes, to capture speech from a target individual, and to reduce capture of background sound and noise originating from nearby locations.

The improved steerable-beam microphone array described in the present application monitors the lobes of the steerable microphone array to identify lobes having large speech content and low noise content. One of the identified lobes is then used to deliver speech to a speech recognition system, as at a self-service kiosk. Figure 7 illustrates one embodiment of the present invention.

Figure 7 illustrates an array of microphones 100, together with lobes L1 - L6. The processing of the signals of microphones M1 and M4 will be taken as representative of the processing of the others. Microphone M1 produces an analog signal S1, and microphone M2 produces an analog signal S2. Those signals are sampled by sample-and-hold circuitry S/H. Dots D represent the samples. Each sample D is digitized by analog-to-digital circuitry A/D, producing a sequence of numbers. Each arrow A represents a number. Each number is stored at an address AD in memory MEM.

Therefore, as thus far described, the system generates a sequence of numbers for each microphone, with each sequence being stored in a separate range of memory MEM. The signals produced, sampled and digitized include speech signals and noise signals.

Beam steering apparatus 200 processes the stored numbers, to generate selected individual lobes L1 - L6 for other apparatus to analyze. The other apparatus includes speech detection apparatus 205, noise detection apparatus 210, and speech recognition apparatus 215. Each apparatus 200, 205, 210, and 215 individually is known in the art, and commercially available.

The basic problem addressed by the present invention is the selection of a lobe which (1) maximizes the speech signal received, and (2) minimizes the noise signal received. The noise of interest is not primarily white noise, but noise from

an artificial source. The frequency components of the noise will not, in general, be equally distributed from zero to infinity. Two examples of the noise in question are (1) a humming air conditioner, and (2) an idling delivery truck. The symbol NC will be used herein to represent this type of noise signal.

Figure 8 is a flow chart illustrating one approach to maximizing signal-to-noise ratio S/NC. In block 300, the lobes L are generated from the data stored in memory MEM in Figure 7, and each is examined. The N lobes carrying the strongest speech signals S are identified. In block 305, the M lobes L carrying the strongest noise signal NC are identified. While these blocks 300 and 305 are represented as separate steps, and in many cases can be executed separately, they can also be executed together.

Identification of the presence of speech signals is well known. For example, speech is discontinuous, while many types of artificial noise, such as the hum of an air conditioner, are continuous and non-pausing. Consequently, the pauses are a feature of speech. Pauses can be detected by, for example, comparing long-term average energy with short-term average energy. In the case background noise, such as originates from an air conditioner unit, the short-term average energy, periodically measured during intervals of a few seconds, will be the same as the long-term average energy, measured over, say 30 seconds. In contrast, for speech, the short-term average energy, similarly measured, but during periods of sound as opposed to silence, will be higher than the long-term average. A primary reason is that the pauses in speech, which contain silence, reduce the long-term average.

Identification of continuous noise is also well known. Two types of continuous noise should be distinguished. If the noise is truly continuous, as in the constant hiss of air flowing through a heating duct, then derivation of a Fourier spectrum can identify the noise as non-speech. In theory at least, a constant, non-

changing, Fourier spectrum will be found. This constant spectrum is not found in speech, and identifies the sound as continuous noise.

In contrast to truly continuous noise, the noise may be continuous, but pulsating, as in an idling gasoline engine. Such noise is continuous, in the sense that it is ongoing, but is also constantly changing, since it is a series of acoustic pulses. Pulses change because they are ON, then OFF, then ON, as it were. Pulsating noise will be characterized by a periodically changing Fourier spectrum, which also distinguishes the noise from speech.

Once blocks 300 and 305 identify the lobes having the highest speech and noise signals, block 310 takes the ratio S/NC for each lobe, and identifies the lobe having the highest ratio. In block 315, that lobe is used to perform speech recognition, by the apparatus 215 in Figure 7.

The processing of blocks 300, 305, and 310 is undertaken by the apparatus 200, 205, 210, and 215 in Figure 7, either individually or collectively. Those apparatus are given access to memory MEM, as indicated by busses B. Those apparatus can also share variables and computation results, as indicated by dashed bus B1.

(6) ISSUES

Whether claims 3 and 4 were properly rejected under 35 U.S.C. §103(a) as being unpatentable over Martino et al. in view of Slyh et al. and Nagata.

Whether claims 5 and 6 were properly rejected under 35 U.S.C. §103(a) as being unpatentable over Martino et al. in view of Nagata.

(7) GROUPING OF CLAIMS

Claims 3 through 6 stand and fall together.

(8) ARGUMENT

Rejection of Claims 3-4 under 35 U.S.C. 103(a)

The rejection of claims 3 and 4 under 35 U.S.C. §103(a) as being unpatentable over Martino in view of Slyh et al. and Nagata is respectfully traversed, as (i) the references, singly or in combination, fail to teach or disclose all limitations of the rejected claims and (ii) there is no suggestion or motivation to combine the references.

Regarding (i), the Final Office Action states "Nagata discloses that all peaks above a threshold are detected as sound sources (Nagata Col. 10, Ln. 4-5)," which, the Official Action maintains, is the equivalent of identifying lobes having a relatively low noise content. Applicant respectfully disagrees. Specifically, "Peaks above threshold may be detected as the sound sources" as taught by Nagata is not equivalent to "identifying lobes having a relatively low noise content," as recited in claims 3 and 4 of the present application. Although the present application teaches a function similar to that taught by Nagata, i.e., that "a minimal level of sound can be established which is considered acceptable," (page 13, lines 3-4), this is not what the Applicant is claiming as the invention. Both the Applicant's excerpt from page 13, line 3-4 and the Nagata excerpt refer to sound, not noise, and both of these excerpts are concerned with the relative intensity of that sound.

The entire sentence from which the Nagata excerpt was taken reads, "By setting a prescribed threshold with reference to an average value of portions other than peak portions on the synthesized (total) sound source power distribution, such as 5 dB, and all peaks above this threshold may be detected as the sound sources, while not detecting any sound source at all when there is no peak above this threshold (Col. 9, Ln. 67 through Col. 10, Ln. 6)." "Above this threshold" refers

back to 5 dB (Col. 10, Ln. 3), which is a measure of the intensity of sound. The measure makes no distinction to type of sound, i.e., speech or noise, as suggested in the Office Action. Furthermore, when "all peaks above this threshold may be detected as a sound source" is read in light of the entire Nagata disclosure, it is obvious that the reference is intended to reduce the calculations necessary to carry out the disclosed method.

Moreover, in reference to element "D," the Final Office Action states "it is obvious to actuate a lobe having both a relatively high speech content and relatively low noise content since one in the art would obviously like to put the prior signal processing to use in a meaningful way." The desire to put the prior signal processing to use in a meaningful way does not make actuating a lobe having both a relatively high speech content and a relatively low noise content obvious. Prior signal processing is essentially captured data or information. Any time a person deliberately causes data or information to be captured that person intends to use that data or information in a meaningful way. This does not make all meaningful uses of that data or information obvious. The same is true in the instant case. While "actuating a lobe having both a relatively high speech content and relatively low noise content" is one possibility among a plethora of meaningful uses that could be conceived, it does not logically follow that this one possibility is obvious. To hold such would be using impermissible hindsight.

Regarding (ii), the Official Action states that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the method of Martino et al. to further include within a kiosk a steerable beam microphone array , having multiple lobes; ii) means for sampling lobes, and A) distinguishing the difference between speech content and noise content from sound signals received by each lobe, B) identifying lobes having a relatively high speech content, C) identifying lobes having a relatively low noise content, and D)

actuating a lobe having both a relatively high speech content and relatively low noise content."

It is well established that "there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. *See* MPEP 2143

Applicant respectfully disagrees with the position taken in the Final Office Action that it would have been obvious to one of ordinary skill in the art at the time of invention to combine the references. Applicant has carefully reviewed the applied references and can find no teachings in the references that support an obviousness rejection. Evidence showing that the suggestion or motivation to modify/combine was actually in the knowledge generally available to one of ordinary skill in the art at the time the present invention was made is respectfully requested.

Thus, as the references, singly or in combination, fail to teach or disclose all limitations of the rejected claims and there is no suggestion or motivation to combine the references, claim 3 and 4 are unobvious and these claims should now be allowed.

Rejection of Claims 5 and 6 under 35 U.S.C. 103(a)

The rejection of claims 5 and 6 under 35 U.S.C. §103(a) as being obvious over Martino in view of Nagata is respectfully traversed. Applicant reasserts that (i) the references, singly or in combination, fail to teach or disclose all limitations of the rejected claims and (ii) there is no suggestion or motivation to combine the references.

In regards to (i), when addressing the argument, in Applicant's December 12, 2003 response, the Final Office Action simply states, "Examiner disagrees,"

never addressing the merits of the Applicant's argument. The Final Official Action cites, as did the December 12, 2003 Office Action, that "Nagata discloses that all peaks on the sound source distribution above a threshold are detected as sound sources (Col. 10, Ln. 4-5)," and that "this is the equivalent of identifying lobes having a relatively low noise content." As explained earlier, this assertion is not support by Nagata.

In fact, "Peaks above threshold may be detected as the sound sources" as taught by Nagata is not equivalent to "identifying lobes having a relatively low noise content," as taught by Applicant. Although the present application teaches a function similar to that taught by Nagata, i.e., that "a minimal level of sound can be established which is considered acceptable," (page 13, lines 3-4), this is not what the Applicant is claiming as the invention. Both the Applicant's excerpt from page 13, line 3-4 and the Nagata excerpt refer to sound, not noise, and both of these excerpts are concerned with the relative intensity of that sound.

The entire sentence from which the Nagata excerpt was taken reads, "By setting a prescribed threshold with reference to an average value of portions other than peak portions on the synthesized (total) sound source power distribution, such as 5 dB, and all peaks above this threshold may be detected as the sound sources, while not detecting any sound source at all when there is no peak above this threshold (Col. 9, Ln. 67 through Col. 10, Ln. 6)." "Above this threshold" refers back to 5 dB (Col. 10, Ln. 3), which is a measure of the intensity of sound. The measure makes no distinction to type of sound, i.e., speech or noise, as suggested in the Office Action. Furthermore, when "all peaks above this threshold may be detected as a sound source" is read in light of the entire Nagata disclosure, it is obvious that the reference is intended to reduce the calculations necessary to carry out the disclosed method.

Moreover, in reference to element "d," the office action states "it is obvious to select a lobe which carries larger speech signals than other lobes and smaller noise signals than other lobes since one in the art would obviously put the prior signal processing to use in a meaningful way in order to enhance speech recognition capabilities." The desire to put the prior signal processing to use in a meaningful way does not make selecting a lobe which carries larger speech signals than other lobes and smaller noise signals than other lobes obvious. Prior signal processing is essentially captured data or information. Any time a person deliberately causes data or information to be captured that person intends to use that data or information in a meaningful way. This does not make all meaningful uses of that data or information obvious. The same is true in the instant case. While "selecting a lobe which carries larger speech signals than other lobes and smaller noise signals than other lobes" is one possibility among a plethora of meaningful uses that could be conceived, it does not logically follow that this one possibility is obvious. To hold such would be using impermissible hindsight.

Regarding (ii), the Official Action states that "it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the method of Martino et al. to further comprise maintaining a beam-steerable microphone array at the self-service kiosk, measuring noise content and speech content of several lobes of the array, and selecting a lobe which carries larger speech signals than other lobes and smaller noise signals than other lobes because one of ordinary skill in the art would recognize that this would provide more accurate speech recognition for suppressing background noise and localizing sound sources effectively."

It is well established that "there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one

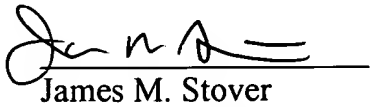
of ordinary skill in the art, to modify the reference or to combine reference teachings. *See* MPEP 2143

Applicant respectfully disagrees with the position taken in the Office Action that it would have been obvious to one of ordinary skill in the art at the time of invention to combine the references. Applicant has carefully reviewed the applied references and can find no teachings in the references that support an obviousness rejection. Should the Examiner insist otherwise, convincing evidence showing that the suggestion or motivation to modify/combine was actually in the knowledge generally available to one of ordinary skill in the art at the time the present invention was made is respectfully requested.

The contention that the references teach each and every element of the claim 5 and 6 is not supported by the references or the Office Action. Nor has the Examiner shown convincing evidence that the suggestion or motivation to modify/combine was actually in the knowledge generally available to one of ordinary skill in the art at the time of invention. Thus, for these reasons claims 5 and 6 of the present application are believed to be patentable over Martino and Nagata.

Review of the present application and claims with consideration of the foregoing comments, and reconsideration of the rejection of claims 3 through 6, are respectfully requested.

Respectfully submitted,



James M. Stover
Reg. No. 32,759

NCR Corporation
1700 South Patterson Blvd., WHQ-4W
Dayton, Ohio 45479-0001
Tel. No. (937) 445-7663
Fax No. (937) 445-6794

(9) APPENDIX

1. (previously presented) Apparatus comprising:

a) a self-service kiosk which dispenses articles, currency, or communication services; and

b) within the kiosk, a steerable-beam microphone array which points a microphone lobe toward a position emanating the highest signal-to-noise ratio, for receiving speech from a customer.

2. (original) System according to claim 1, wherein the system further comprises speech recognition apparatus for recognizing said speech.

3. (previously presented) Apparatus comprising:

a) a self-service kiosk which dispenses articles, currency, or communication services; and

b) within the kiosk,

i) a steerable beam microphone array, having multiple lobes;

ii) means for sampling lobes, and

A) distinguishing the difference between speech content and noise content from sound signals received by each lobe,

B) identifying lobes having a relatively high speech content,

C) identifying lobes having a relatively low noise content, and

D) actuating a lobe having both a relatively high speech content and relatively low noise content.

4. (original) Apparatus according to claim 3, and further comprising:

c) speech recognition means for recognizing speech contained in the lobe actuated.

5. (original) A method, comprising the following steps:

a) maintaining a self-service kiosk which dispenses articles, currency, or communication services;

b) maintaining a beam-steerable microphone array at the self-service kiosk;

c) measuring noise content and speech content of several lobes of the array; and

d) selecting a lobe which carries

i) larger speech signals than other lobes and

ii) smaller noise signals than other lobes.

6. (previously presented) Method according to claim 5, and further comprising the step of:

e) receiving signals from the lobe selected, and performing speech recognition on the data.